

## Olive: A Digital Archive for Executable Content

Gloriana St. Clair, Dan Ryan, Mahadev Satyanarayanan(Satya), Vasanth Bala, Erika Linke  
Coalition for Networked Information, December 12-13, 2011

**Purpose:** “Today, an increasing fraction of the world’s intellectual output is in the form of executable content. These include simulation models, tutoring systems, expert systems, data visualization tools, and so on. Even content that could be static (such as a company’s product information Web page) is often dynamically generated through execution of code that customizes the content and appearance at runtime... [Olive seeks] to freeze and precisely reproduce the execution state that produced this dynamic content.”<sup>1</sup>

### History:

2008 – Satya and Vas conceive of Olive, IBM reviewers suggest library application

2009 – Satya contemplates long term preservation and opens communication with Gloriana St. Clair to discuss past efforts and current preservation responsibilities of the library

2010 – Dave McQueeney at IBM hosts invitational workshop arranged by Satya, Vas, Joel Smith (CMU CIO), Gloriana, and Erika

2011 – Workshop held at IBM Watson Labs (in the wake of Watson's Jeopardy win)

*Attendees: Gloriana St. Clair (CMU), Mahadev Satyanarayanan (CMU), Erika Linke (CMU), Vasanth Bala (IBM), Dianne Fodell (IBM), Giovanni Pacifici (IBM), Clifford Lynch (CNI), Chris Shillum (Elsevier), Mike Handy (Library of Congress), Claire Cocco (OCLC), Mike Lesk (ACM), Patrick Dreher (RenCI), Timothy Dilauro (Johns Hopkins), Kevin Guthrie (Ithaka), Paul Jones (ibiblio), Vijay Kumar (MIT), Naomi Nelson (Duke), Peng Ning (NCSU), Mladen Vouk (NCSU), Joan Smith (Emory), Michael Witt (Purdue)*

Feedback:

- Group excited to see demonstration of old software actually running in Olive
- Softwares should run as described on original packaging
- Many IP issues to consider going forward
- Extensibility concerns – Many-to-one relationship for executables → parcels
- Metadata tracking and format questions
- Sustainability

Conclusion: Articulate the next logical steps

- Satya and Vas to pursue funding for bootstrapping
- Gloriana and Erika to pursue funding for the mounting use cases

### Software & Ecosystem:

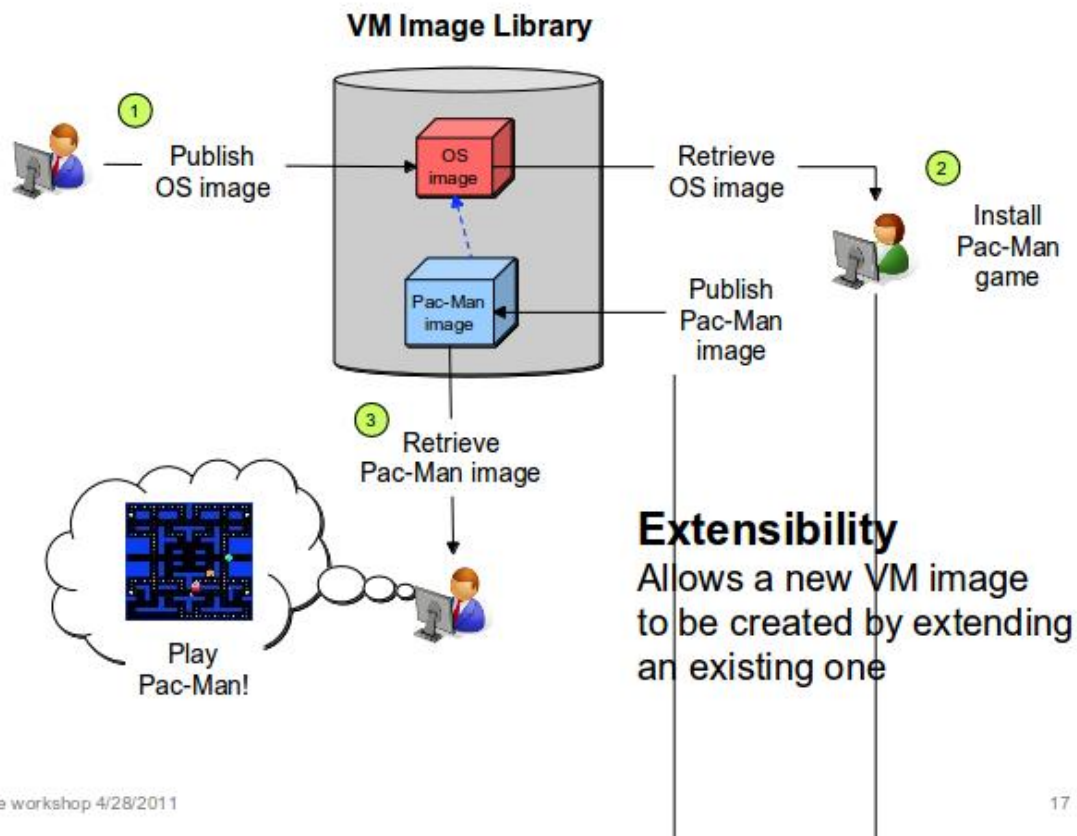
- Software developed by Satya and Vas enables end users to stream virtual machines remotely
  - Fills a gap in current computing model

Location	Remote Execution	Local Execution
Remote Storage	Cloud	Olive
Local Storage	Grid Computing	Traditional PC

- Users can modify virtual machine (install or use software) and save only the changes
- Base software remains unmodified
- Virtual machines divorced from individual software packages which can be added as needed
- Not reliant on (impractical) hardware preservation

<sup>1</sup> Satyanarayanan, M. et. al., *Collaborating With Executable Content Across Space and Time*, Proceedings of the 7<sup>th</sup> International Conference on Collaborative Computing: Networking, Applications and Worksharing, October 2011

# Technology and Infrastructure requirements



17

## Challenges:

- Implementation
- Ease of use
- Sustainable business model
- Intellectual Property
- Scalability
- Unknown unknowns

## Evolving IMLS Proposal:

- Implement Olive for preservation of real world software to identify unknown unknowns
  - Educational software: Dave Yaron's *ChemLab* digital chemistry learning resource from Carnegie Mellon University
  - Games: Preserving Virtual Worlds team (previous IMLS grant) to help select games and implement best practices in preservation of executable content
  - Scholarly Articles: Special issue of an Elsevier journal containing executable content with support from Elsevier research team